



PREDICTION OF TURBIDITY IN TIGRIS RIVER USING ARTIFICIAL NEURAL NETWORKS

Rafa H. Al-Suhaili Waleed M.S.Kassim Yousif M.Yousif
Department of Environmental Engineering / College of Engineering /
Baghdad University.

ABSTRACT

Over the past two decades, there has been an increased interest in a new class of computational intelligence systems known as Artificial Neural Networks (ANNs). In this work, (ANNs) technique was applied in an attempt to predict the turbidity at intake of Al-Wathba water treatment plant (WTP) in Baghdad. This prediction is useful in the planning, evaluation, management, and operation of such plants, which may produce water of better quality. The available records from (1991-2000) were used for predicting turbidity in Tigris River, based on monthly maximum values of the water quality parameters near intakes of the water treatment plants. Multi-layer perceptron trainings using the back-propagation algorithm were used in this work. The feasibility of ANNs technique for modeling this water quality parameter was investigated. A number of issues in relation to ANNs construction such as the effect of ANNs geometry and internal parameters on the performance of ANNs model were investigated. It was found that ANNs have the ability to predict the Turbidity at Al-Wathba WTP with a good degree of accuracy (the coefficient of determination (R^2) was 0.9687). The ANNs model developed to study the impact of the internal network parameters on model performance indicate that ANNs performance was relatively insensitive to the number of hidden layer nodes, momentum term, and learning rate.

INTRODUCTION

Water quality control and management have attached increasing attention from developing countries. The aspects of environmental protection are becoming a major obstacle for further and sustainable economic and social development. The use of raw water quality data cannot be overlooked, since such data have considerable effect on the calculation of needed chemical, proper management, treatment, and assessing the potentialities of the use of furnished water supplies for different purposes, (**Barzanjy, 2007**).

Traditionally, there have been two main philosophical approaches for surface water quality modeling. Process based models, which consider the underlying physical processes directly, and statistical models, which determine relationships based on historical data sets. Recently, artificial neural networks (ANNs) have emerged as alternatives to traditional statistical models in a variety of fields, including water quality management, (Maier and Dandy, 2000).

ANNs are one of the modeling techniques that attempt to simulate the operation of the human brain and nervous systems. ANNs learn by 'example' in which an actual measured set of input variables and the corresponding outputs are presented to determine the rules that govern the relationship between variables. Consequently, ANNs are well suited to modeling complex problems where the relationship between the variables is unknown and when non-linearity is suspected. Although, the concept of artificial neurons was first introduced in 1943 by McCulloch and Pitts, research into applications of ANNs has blossomed since the introduction of the back-propagation-training algorithm for feed forward ANNs in 1986. ANNs may thus be considered a relatively new tool in the field of prediction and forecasting, (Shahin et. al. 2003).

AREA AND OBJECTIVE OF THE STUDY

The Tigris River has a large importance in the present and in the future. This is because of the detrimental effect of pollutants resulting from human activities, industrial wastes, sewage wastes, and harmful effect of increasing drainage waters coming from agricultural lands upstream coupled with the decrease in its discharge, (Abu-Hamdeh, 2000). So it has become necessary to make detailed studies and researches to evaluate the suitability of river water for different purposes in a selected site of the river.

Baghdad city was chosen for this study. Artificial neural network was applied to predict the turbidity at the intakes of the existing water treatment plants in Baghdad. The running conventional water treatment plants are (Karkh, Sharq Dijlah, Karama, Wathba, Qadisiya, Dora, and Rasheed) as shown in Fig. (1).

The records from (1991-2000) were used for predicting turbidity at Al-Wathba WTP of monthly maximum values for the water quality parameters near intakes of the water treatment plants.

The main objective of this study is the evaluation and management for some of the water quality parameters at intakes of the water treatment plants in Baghdad by using ANNs technique to predict the turbidity at intake of Al-Wathba water treatment plant in Baghdad.

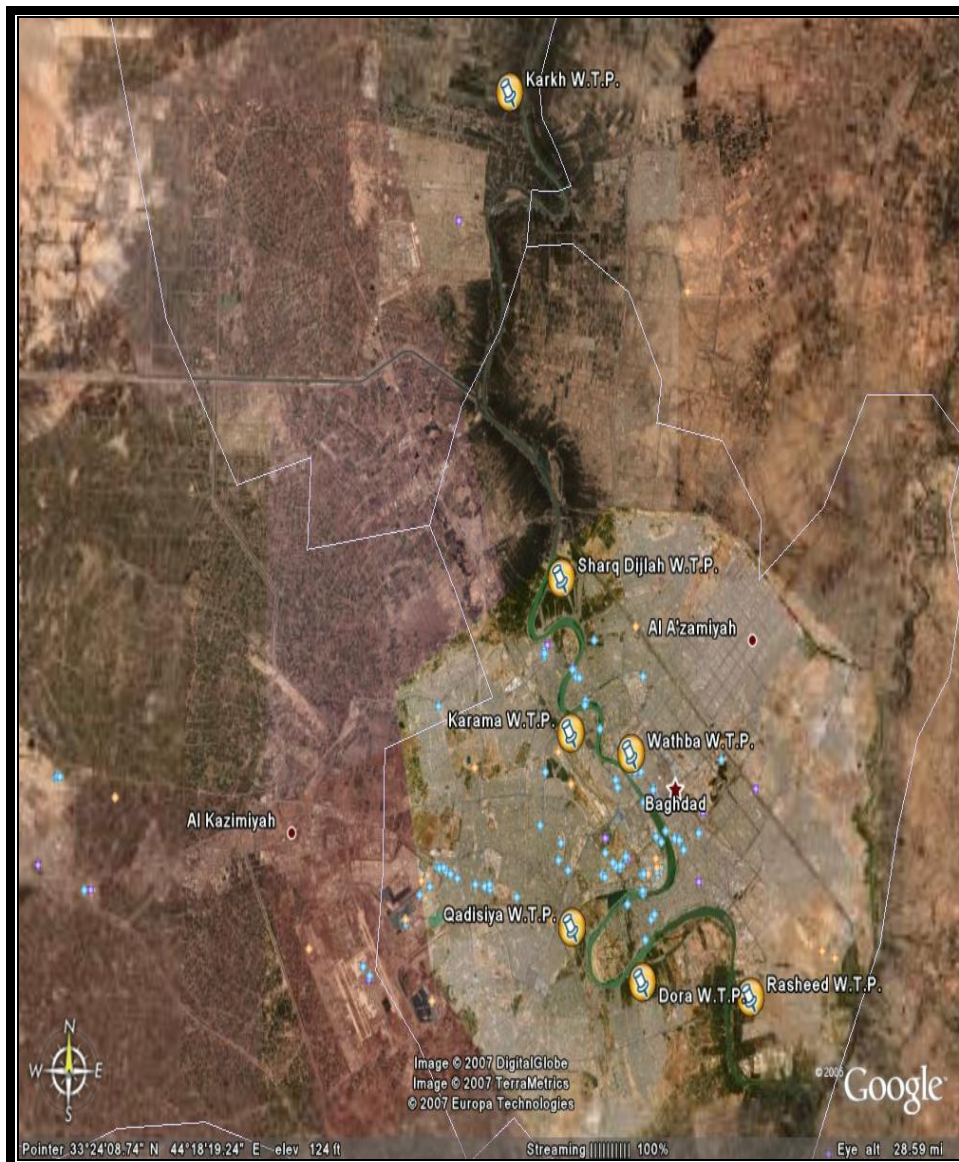


Fig. (1) View of the Study Area

DETERMINATION OF ARTIFICIAL NEURAL NETWORKS MODELS

Artificial neural networks (ANNs) models need to be in a systematic manner to improve its performance. Such an approach needs to address major factors such as, determination of model inputs, data division and pre-processing, determination of model architecture, model optimization (training), and model validation, (**Maier and Dandy, 2000**). These factors are explained and discussed below.

A PC-based commercial software system was used, called *Neuframe* program (Version 4.0) (Neosciences 2000, Neurosciences Crop., Southampton, Hampshire, U.K.). The optimal network architecture was determined by trail-and-error.

Determination of Model Inputs and Outputs

The selection of the model input variables that have the most significant impact on the model performance is an important step in developing ANNs models. Presenting as large a number of input variables as possible to ANNs models usually increases network size, resulting in a decrease in processing speed and a reduction in the efficiency of the network, (**Zaheer and Bai, 2003**).

Different approaches have been suggested to assist the selection of input variables. The approach that was adopted in this research based on priori knowledge, the appropriate input variables can be selected. This approach is usually utilized in the field of civil engineering, (**Maier and Dandy, 2000**).

The inputs of this model are:

- Flow of the river, Q.
 - The Turbidity at Al-Karkh WTP, Tur-1.
 - The Turbidity at Sharq Dijlah WTP, Tur-2.
 - The Turbidity at Al-Karama WTP, Tur-3.
 - Suspended Solid at Al-Karkh WTP, S.S-1.
 - Suspended Solid at Sharq Dijlah WTP, S.S-2.
 - Suspended Solid at Al-Karama WTP, S.S-3.
 - The Distance between Al-Karkh and Al-Wathba WTP, D-1,4.
 - The Distance between Sharq Dijlah and Al-Wathba WTP, D-2,4.
 - The Distance between Al-Karama and Al-Wathba WTP, D-3,4.
- The output of the model is the turbidity at Al-Wathba WTP, Tur-4.

DATA DIVISION AND PRE-PROCESSING

It is a common practice to divide the available data into three sets, training, testing, and validation. The training set is used to adjust the connection weights of the neural network. The testing set is used to check the performance of the network at various stages of learning, and training is stopped once the error in the testing set increases. The validation set is used to evaluate the performance of the model once training has been successfully accomplished. The way data are divided can have a significant effect on model performance, (**Al-Janabi, 2006**), and trail-and-error process was used to select the best division.

Once the available data have been divided into their subsets (i.e. training, testing and validation), it is important to pre-process the data in a suitable form before they applied to the ANNs. Data pre-processing is necessary to ensure all variables receive equal attention during the training process. Pre-processing can be in the form of data scaling, normalization and transformation. Thus, the logarithm of inputs and outputs of this model were taken (except the distances) before proceeding forward in the next steps.

DETERMINATION OF MODEL ARCHITECTURE

Determining the network architecture is one of the most important and difficult tasks in ANNs model development. It requires the selection of the optimum number of layers and the number of nodes in each of these layers. There is no unified theory for the determination of an optimal ANNs architecture. It is generally achieved by fixing the number of layers and choosing the number of nodes in each layer.

There are always two layers representing the input and output variables in any neural networks. It has been shown that one hidden layer is sufficient to approximate any continuous function provided that sufficient connection weights are given, (Al-Neami, 2006). The number of nodes in the input and output layers are restricted by the number of model inputs and outputs, respectively. There is no direct and precise way of determining the best number of nodes in each hidden layer. A trial-and-error procedure, which is generally used in civil engineering to determine the number and connectivity of the hidden layer nodes, can be used, (Shahin et. al, 2002).

(Resop, 2006) suggested that the upper limit of the number of hidden nodes in a single layer network may be taken as $(2I+1)$, where I is the number of inputs. The best approach found by (Nawari et. al, 1999) was to start with a small number of nodes and to slightly increase the number until no significant improvement in model performance is achieved.

MODEL OPTIMIZATION (TRAINING)

As mentioned previously, the process of optimizing the connection weights is known as 'training' or 'learning'. This is equivalent to the parameter estimation phase in conventional statistical models. The aim is to find a global solution to what is typically a highly non-linear optimization problem. The method most commonly used for finding the optimum weight combination of feed-forward neural networks is the back-propagation algorithm, which is based on first-order gradient descent, (Jun, 2002). Ultimately, the model performance criteria, which are problem specific, will dictate which training algorithm is most appropriate. If training speed is not a major concern, there is no reason why the back-propagation algorithm cannot be used successfully.

MODEL VALIDATION

Once the training phase of the model has been successfully accomplished, the performance of the trained model should be validated. The purpose of the model validation phase is to ensure that the model has the ability to generalize within the limits set by the training data in a robust fashion, rather than simply having memorized the input-output relationships that are contained in the training data. The approach that is generally adopted in the research to achieve this is to test the performance of trained ANNs on an independent validation set, which has not been used as part of the model building process. If such performance is adequate, the model is deemed to be able to generalize and is considered to be robust. The training error and the testing error (carried out by *Neuframe* software) and the coefficient of correlation of validation set (r) are the main criteria that are often used to evaluate the prediction performance of ANNs models.

RESULTS AND DISCUSSION

The effect of data subsets divisions on performance of ANNs was investigated. Trial-and-error process was used to select the best division, the network that performs best with respect to testing error was used in this work. Using the default parameters of the software, a number of networks with different divisions were developed and the results are shown graphically in Fig. (2). It can be seen from Fig. (2) that the best division is 60 % for training set, 25 % for testing set, and 15 % for validation set, according to lowest testing error. Thus, this division was adopted in this model.

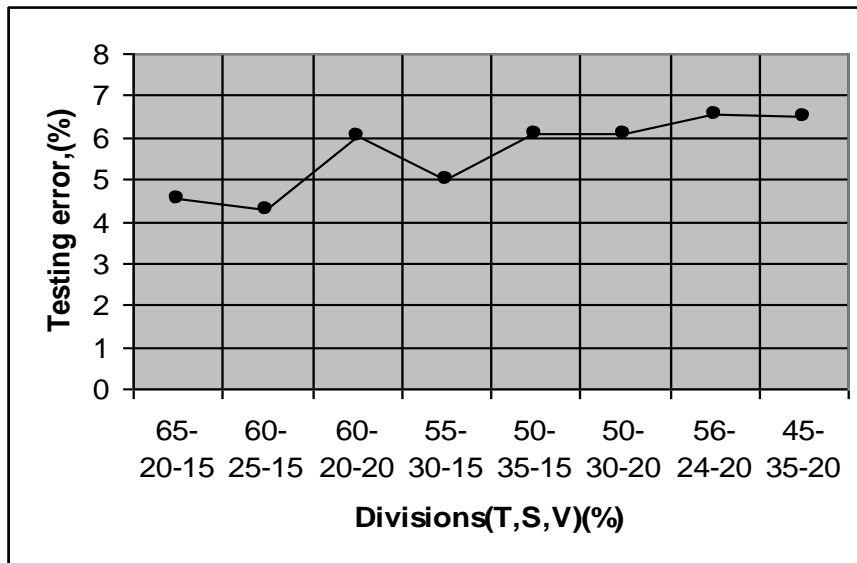


Fig. (2) Effect of data divisions on performance of ANNs

The effect of the number of hidden nodes on ANNs performance was investigated. A number of trials were carried out using the default parameters of the software used with one hidden layer and start with one hidden node and then slightly increasing the number of the nodes until no significant improvement in the model performance, was gained. a number of networks with different numbers of hidden layer nodes were developed and the results are shown graphically in Fig. (3). It can be seen from Fig. (3) that there are slightly differences in the testing error after 8 nodes. Therefore, the process was stopped at 10 nodes where no significant improvement in model performance, was found. Fig. (3) shows that the network with 3 hidden layer nodes has the lowest prediction error for testing test. However, it is believed that the network with 1 hidden layer node is considered optimal, as its prediction error is not far from the network with 3 hidden layer nodes coupled with smaller number of connection weights. Therefore, 1 hidden layer node was chosen in this model.

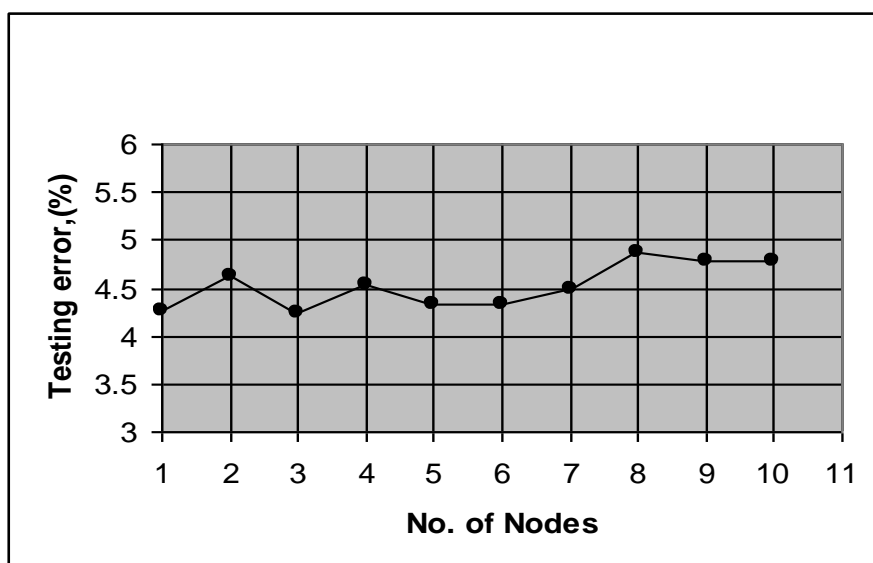


Fig. (3) Performance of ANNs model with different hidden layer nodes

The effect of the internal parameters controlling the back-propagation algorithm (i.e. momentum term and learning rate) on the model performance was investigated for the model with 1 hidden layer node. The effect of the momentum term on model performance is shown graphically in Fig (4). It can be seen that the performance of the ANNs model is relatively insensitive to the variation of the momentum term, particularly in the range 0.01 to 0.60. Then the testing errors slightly decrease at the range 0.80 to 0.95. Thus, the obtained optimum value for momentum term is 0.80, which have the lowest values of testing error and training error, hence it was used in this model.

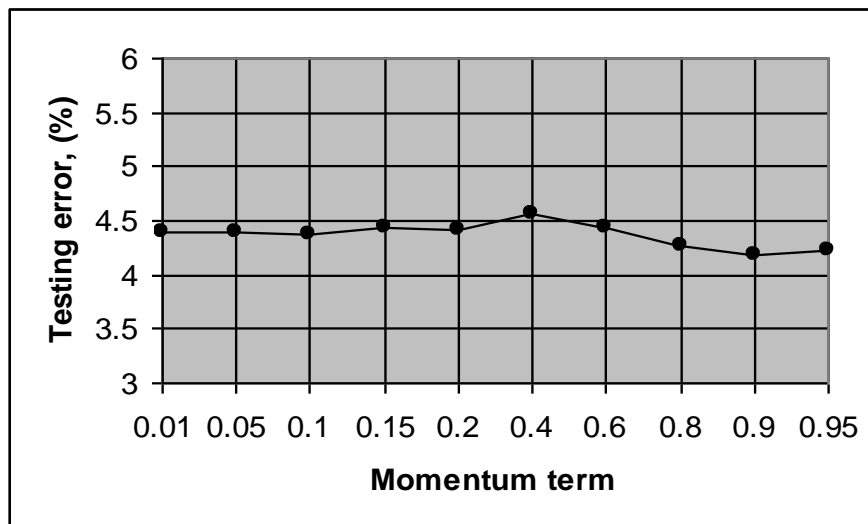


Fig. (4) Effect of various momentum term on ANNs performance

The effect of the learning rate on the model performance is shown graphically in Fig. (5). It can be seen that the performance of the ANNs model is relatively insensitive to the variation of the learning rate. The testing errors are slightly decreased at the range 0.10 to 0.40, while it is slightly increase at the range 0.60 to 0.99. Thus, the obtained optimum value for learning rate is 0.40, which has the lowest value of testing error, hence it was used in this model.

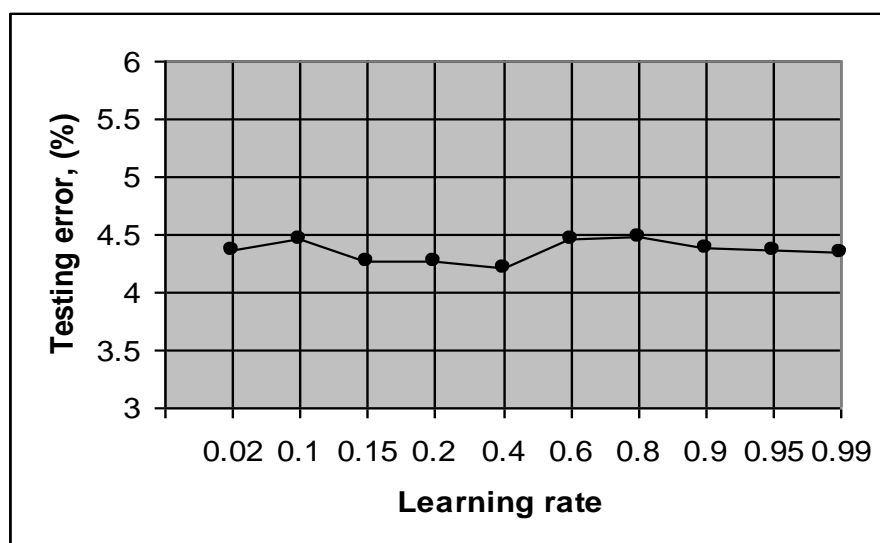


Fig. (5) Effect of various learning rate on ANNs performance

In an attempt to identify which of the input variables have the most significant impact on the

output predictions, a sensitivity analysis was carried out on the ANNs. A simple and innovative technique proposed by (Garson, 1991) was used to interpret the relative importance of the input variables by examining the connection weights of the trained network. The results indicate that the turbidity and suspended solid at Al-Karama WTP have the most significant effect on the predicted the turbidity at Al-Wathba WTP with a relative importance of 16.4 and 16.343 % respectively. The results also indicate that the turbidity and suspended solid upstream have a moderate impact on prediction, while the flow of the river and the distances between those water treatment plants have the smallest impact on the prediction, as shown in Fig. (6).

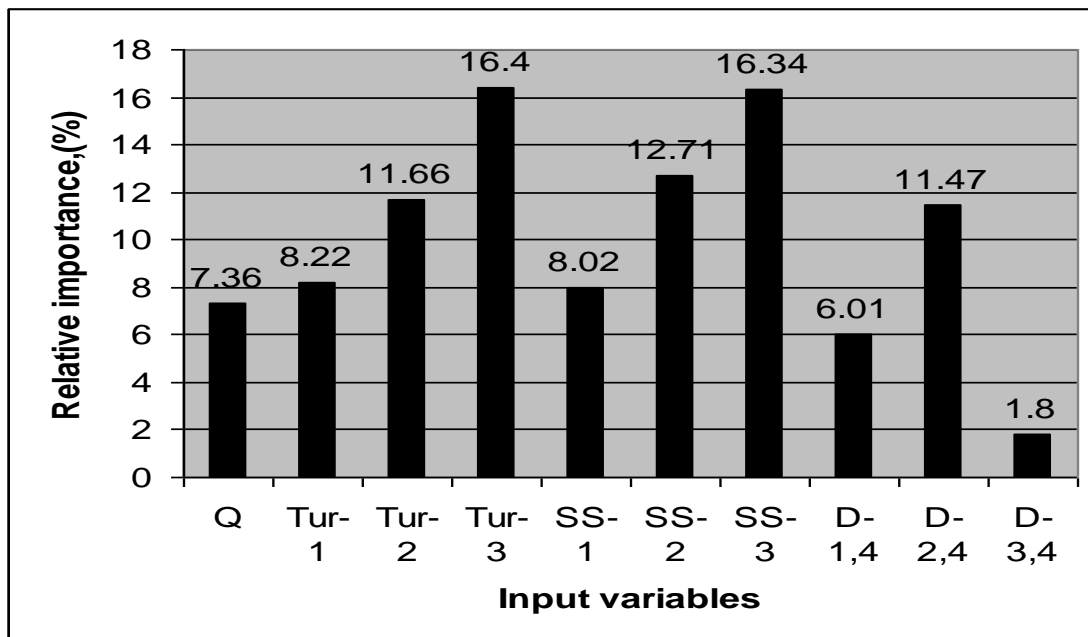


Fig. (6) Relative importance of the input variables

ANNS MODEL EQUATION

The small number of connection weights obtained by **Neuframe** for the optimal ANNs model enables the network to be translated into relatively simple formula. To demonstrate this, the structure of the ANNs model is shown in Fig. (7).

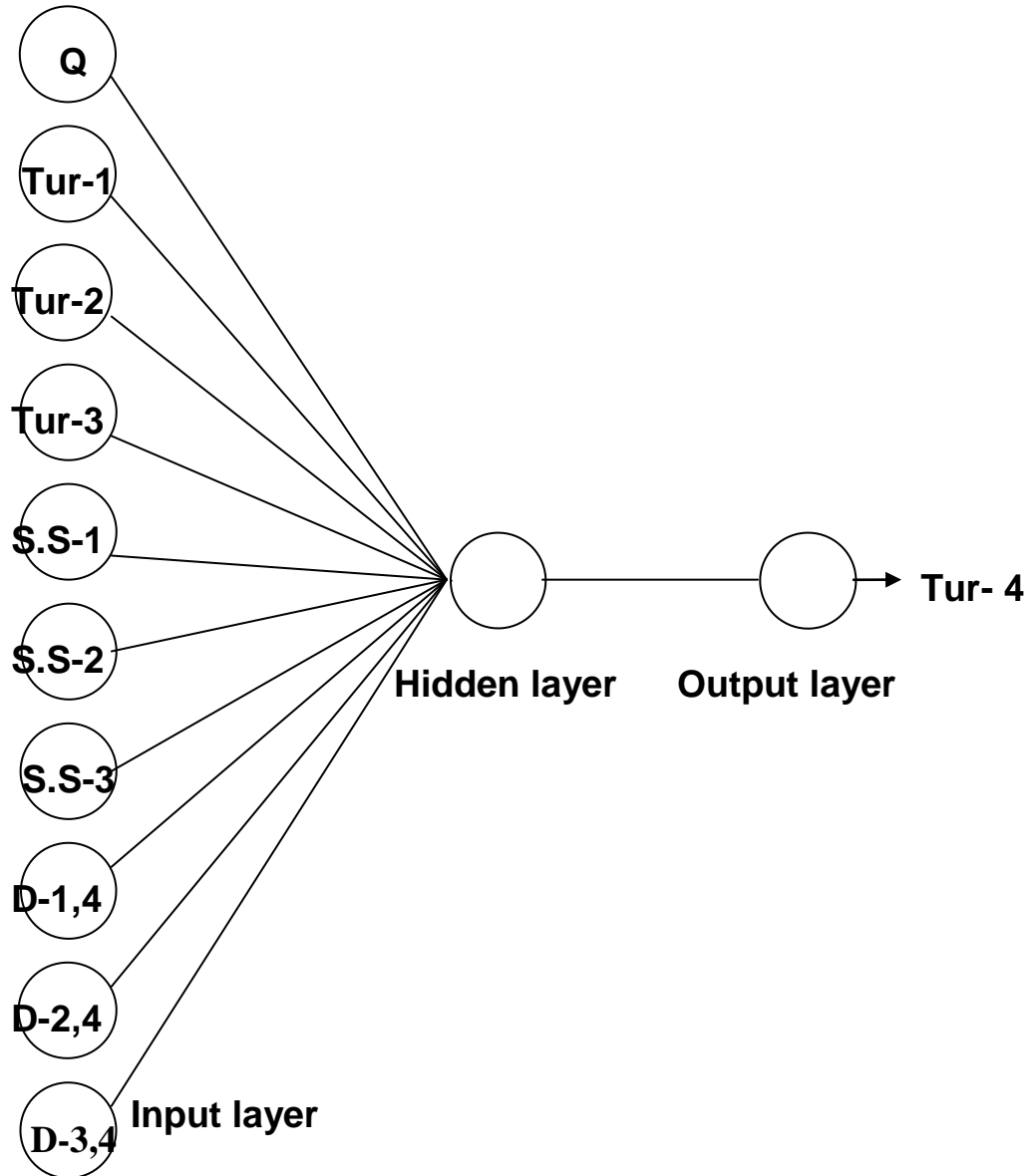


Fig. (7) Structure of the ANNs optimal model

The derived formula is as follows:

$$Tur - 4 = \frac{2.699}{1 + e^{(-1.7+5.595 \tanh x)}} + 1.255 \dots \dots \dots (1)$$

Where:

$$x = 1.233 + 0.735Q - 0.139Tur_1 - 0.307Tur_2 - 0.22Tur_3 - 0.131S.S_1 - 0.253S.S_2 - 0.339S.S_3 \dots \dots \dots (2)$$

VALIDITY OF THE ANNS MODEL

To assess the validity of the ANNs model for the turbidity at Al-Wathba WTP (Tur-4), the predicted values of Tur-4 are plotted against the measured (observed) values of Tur-4 for validation data set, as shown in Fig. (8). It is clear from Fig. (8), the generalization capability of ANNs techniques using the validation data set. The coefficient of determination (R^2) is (96.87 %),

therefore it can be concluded that ANNs model show very good agreement with the actual measurements.

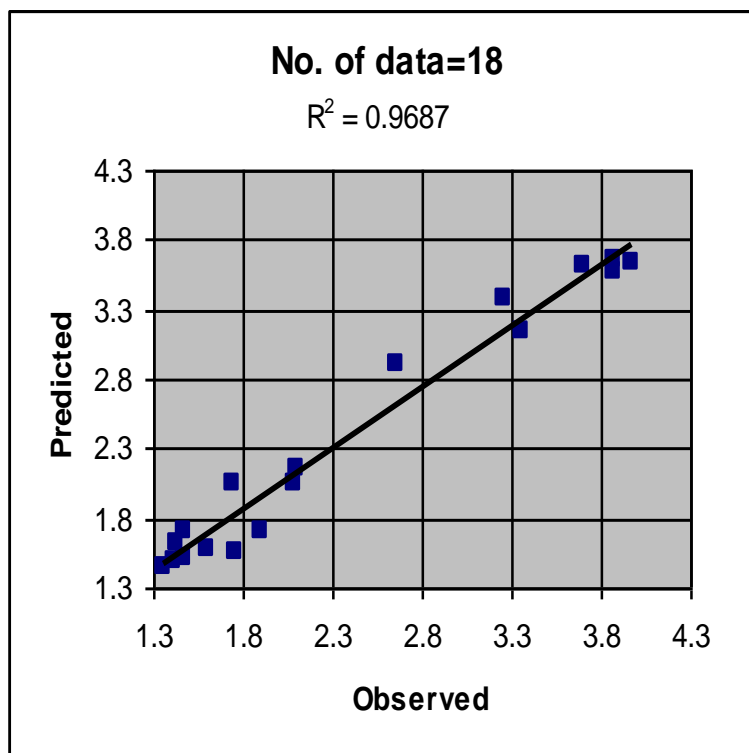


Fig. (8) Comparison of predicted and measured Tur-4 for validation data set

CONCLUSIONS

The results obtained from this work had yielded the following conclusions.

- ANNs have the ability to predict (Tur-4), with a very good degree of accuracy within the range of data used for developing ANNs models.
- The ANNs models developed to study the impact of the internal network parameters on model performance indicate that ANNs performance is insensitive to the number of hidden layer nodes, momentum terms, and learning rate (the optimal network usually achieved by trail and error process).
- The sensitivity analysis indicated that the parameters at Al-Karama WTP have the most significant effect on the prediction of the parameters at Al-Wathba WTP. The results also indicated that the parameters at further upstream of the river have a moderate impact on the prediction, while the flow of the river and the distances between those WTP have the smallest impact on the prediction.
- ANNs models could be translated into simple and practical formula from which (Tur-4) may be calculated.

REFERENCES

- Abu-Hamdeh, M.R.M., (2000), *Study of Tigris Water Quality and Treated Water at the Water Treatment Plants for Baghdad City*, M.Sc. Thesis, Environmental Eng. Dept. University of Baghdad.
- Al-Janabi, K.R.M., (2006), *Laboratory Leaching Process Modeling in Gypseous Soils Using Artificial Neural Network (ANN)*, Ph.D. Thesis, Building and Construction Eng. Dept. University of Technology.
- Al-Neami, M.A.M., (2006), *Evaluation of Delayed Compression of Gypseous Soils with Emphasis on Neural Network Approach*, Ph.D. Thesis, Building and Construction Eng. Dept. University of Technology.
- Barazanji, S.J.I., (2007), *Short and Long Term Forecasting of Water Quality Parameters in Baghdad*, M.Sc. Thesis, Environmental Eng. Dept. University of Baghdad.
- Garson, G. D. (1991), "Interpreting Neural-Network Connection Weights." *AI Expert* 6(7), 47-51.
- Jun, H., (2002), *Application of Artificial Neural Networks for Flood Warning Systems*, Ph.D. - Thesis, Dept. of Civil and Environmental. Eng. University of North Carolina.
- Maier, H.R., and Dandy, G.C., (2000), *Application of Artificial Neural Networks to Forecasting of Surface Water Quality Variables: Issues, Application and Challenges*, Artificial Neural Networks in Hydrology, R.S. Govindaraju, and A.R. Rao, eds., Kluwer, Dordrecht, Netherlands, 287-309.
- Nawari, N.O, Liarg, R. and Nussairat, J., (1999), "Artificial Intelligence Techniques for the Design and Analysis of Deep Foundations", *EJGE*, Vol.4.
- Resop, J.P., (2006), *A Comparison of Artificial Neural Networks and Statistical Regression with Biological Resources Applications*, M.Sc. Thesis, Faculty of the Graduate School of the University of Maryland, Collage Park.
- Shahin, M.A, Jaska, M.B. and Maier, H.R., (2002), "Predicting Settlement of Shallow Foundations Using Neural Networks", *Journal of Geotechnical and Geoenvironmental Engineering*, ASCE, Vol.128, No.9, pp. 785-793.
- Shahin, M.A., (2003), *Use of Artificial Neural Networks for Predicting Settlement of Shallow Foundations on Cohesionless Soils*, Ph.D. Thesis, Department of Civil and Environmental Eng., University of Adelaide.
- Shahin, M.A, Jaska, M.B. and Maier, H.R., (2003), "Application of Artificial Neural Networks in Foundation Engineering", *Australian Geomechanics*.
- Zaheer, I., and Bai, C., (2003), *Application of Artificial Neural Networks for Water Quality Management*, *Lowland Technology International*, Vol. 5, No. 2, 10-15.